|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

**ANS:** The probability of getting two heads and one tail on tossing three coins at

once is equal to 3/8.

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

**ANS:**  The sum is equal to 1 is zero

1. Less than or equal to 4

**ANS:** Less than an equal to 4 = 1/6

1. Sum is divisible by 2 and 3

**ANS:** Sum is divisible by 2 and 3 = 1/6

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

**ANS:** 10/21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

**ANS:**  Expected number of candies for a randomly selected child = 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Please check - Q7**

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

**ANS:** 145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Please check - Q9\_a&b (1)**

**Q210) Draw inferences about the following boxplot & histogram**



**Please check Q10**

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

**ANS:** As sample size is greater than 30 and STD is known we will go with Z value.

Formula for confidence interval is (mean +/- (Z \*STD/root of n))

94% confidence interval

100-94 = 6

6/2 = 3

94 + 3 = 97

Look for 0.97 value in Z table = 1.89

Conf interval for 94%= 200+/- (1.89\*30/root of 2000) = 200+/- 1.2679

So Conf interval for 94% is (198.73, 201.2679)

OR Using Python = stats.norm.interval(0.94,200,30/np.sqrt(2000))

Same way found the conf interval for 98%

Conf interval for 98%= 200+/- 2.33\*30/root of 2000) = 200+/- 1.563

So Conf interval for 98% is (198.437, 201.563)

OR Using Python = stats.norm.interval(0.98,200,30/np.sqrt(2000))

Conf interval for 96%= 200+/- 2.06\*30/root of 2000) = 200+/- 1.3819

So Conf interval for 96% is (198.6181, 201.3819)

OR Using Python = stats.norm.interval(0.96,200,30/np.sqrt(2000))

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.
2. What can we say about the student marks?

**ANS:** Mean – 41

Median – 40.5

Variance - 25.52941

Standard Deviation - 5.052664

**Please check – A1Q12**

Q13) What is the nature of skewness when mean, median of data are equal?

**ANS**: The distribution has zero skewness.

Q14) What is the nature of skewness when mean > median ?

**ANS**: If the mean is greater than the median, the distribution is positively skewed.

Q15) What is the nature of skewness when median > mean?

**ANS:** If the mean is less than the median, the distribution is negatively skewed.

Q16) What does positive kurtosis value indicates for a data ?

**ANS:** Positive values of kurtosis indicate that a distribution is peaked and possess

thick tails.

Q17) What does negative kurtosis value indicates for a data?

**ANS:** Negative values of kurtosis indicate that a distribution is flat and has thin

Tails.

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

**ANS:**  A box plot shows the distribution of data. It is useful in visualizing skewness in data.

What is nature of skewness of the data?

**ANS:**  The nature of the skewness of the data is negatively skewed because the median is wider to the lower quartile.

What will be the IQR of the data (approximately)?

**ANS:** The IQR of the data is approximately 8.

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**ANS:** The data in first boxplot is less distributed and closer to the mean. The data

in the second boxplot is much distributed and deviated from mean. Both

have the same mean*.*

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

**ANS:** a. P = 34/81

b. P = 61/81

c. P = (76-7)/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS:** The MPG of Cars follows Normal Distribution.

1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**ANS:** The Adipose Tissue (AT) and Waist Circumference (Waist) from wc-

atdata set NOT follows Normal Distribution.

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Please check - Z\_score Q22**

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

**ANS:** Df = 25-1 = 24

stats.t.ppf(0.95,24) = 1.71

stats.t.ppf(0.96,24) = 1.828

stats.t.ppf(0.99,24) = 2.49

Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedomt6

**ANS:** pt(-0.471,17)

0.321814

The Probability of the bulbs lasting less than 260 days on average of life

0.321814.

**Please check – Q24**